Whole-genome sequencing reveals Listeria monocytogenes diversity and allows identification of long-term persistent strains in Brazil

Anderson Camargo, Alexandra Moura, Johannetsy Avillan, Nicole Herman, Adelle Mcfarland, Srinand Sreevatsan, Douglas Call, Joshua Woodward, Marc Lecuit, Luís Nero

Whole-genome sequencing reveals *Listeria monocytogenes* diversity and allows identification of long-term persistent strains in Brazil

Anderson C. Camargo[1#], Alexandra Moura[2,3#], Johannetsy Avillan[4], Nicole Herman[5], Adelle P. McFarland[6§], Srinand Sreevatsan[7], Douglas R. Call[4], Joshua J. Woodward[6], Marc Lecuit[2,3,8], Luís A. Nero[1]*

[1]Departamento de Veterinária, Universidade Federal de Viçosa, Viçosa, MG, Brazil; [2]Institut Pasteur, Biology of Infection Unit, Paris, France; [3]Inserm U1117, Paris, France; [4]Paul G. Allen School for Global Animal Health, Washington State University, Pullman, Washington State, USA;

[5]Veterinary Population Medicine Department, University of Minnesota, St. Paul, Minnesota, USA; [6]Department of Microbiology, University of Washington, Seattle, Washington State, USA; [7]Department of Pathobiology and Diagnostic Investigation, Michigan State University, East Lansing, Michigan, USA; [8]Université de Paris, Department of Infectious Diseases and Tropical Medicine, Necker-Enfants Malades University Hospital, APHP, Institut Imagine, Paris, France.

[#]These authors contributed equally to this article

§Present address: Department of Pathology & Immunology, Washington University, St. Louis, Missouri, USA

Address for correspondence: nero@ufv.br, mlecuit@pasteur.fr

Summary

Advances in whole-genome sequencing (WGS) technologies have document genetic diversity and epidemiology of the major foodborne pathogen *Listeria monocytogenes* (*Lm*) in Europe and North America, but data concerning South America is scarce. Here we examined the population structure and genetic diversity of this major foodborne pathogen collected in Brazil. Based on core genome multilocus sequence typing (cgMLST), isolates from lineages I (n=22; 63%) and II (n=13; 37%) were distributed into 10 different sublineages (SLs) and represented 31 new cgMLST types (CTs). The most prevalent SLs were SL9 (n=9; 26%), SL3 (n=6; 17%), SL2 and SL218 (n=5; 14%). Isolates belonging to CTs L2-SL9-ST9-CT4420 and L1-SL315-ST520-CT4429 were collected 3 and 9 years apart, respectively, revealing long-term persistence of *Lm* in Brazil. Genetic elements associated with stress survival were present in 60% of isolates (57% SSI-1 and 3% SSI-2). Pathogenic islands were present in 100% (LIPI-1), 43% (LIPI-3), and 6% (LIPI-4) of the isolates. Mutations leading to premature stop codons were detected in the *prfA* and *inlA* virulence genes. This study is an important contribution to understanding the genomic diversity and epidemiology of *Lm* in South America. In addition, the results highlight the importance of using WGS to reveal *Lm* long-term persistence.

Key words: *Listeria monocytogenes*, whole genome sequencing, molecular typing, virulence, resistance, persistence

Introduction

The Gram-positive bacterium *Listeria monocytogenes* (*Lm*) is an important foodborne pathogen that is able to survive in stressful environmental conditions and grow in refrigerated foods. The consumption of contaminated food has been linked with both epidemic and sporadic listeriosis. Despite the low overall incidence of listeriosis, the disease is associated with high hospitalization and mortality rates (Orsi et al., 2011; de Noordhout et al., 2014; Lomonaco et al., 2015; Charlier et al., 2017). Invasive forms of the disease are particularly dangerous for immunocompromised individuals and pregnant woman, leading to sepsis, meningitis, encephalitis, or abortion (Silk et al., 2012; de Noordhout et al., 2014; Charlier et al., 2017).

Lm population structure is relatively clonal, divided into four phylogenetic lineages (I, II, III, and IV) (Orsi et al., 2011) and multiple clonal complexes (CCs, defined on the basis of multilocus sequence typing [MLST]) (Ragon et al., 2008) or sublineages (SLs, defined on the basis of core genome MLST) (Moura et al., 2016). Major CCs and SLs are distributed globally (Chenal-Francisque et al., 2011; Haase et al., 2014; Moura et al., 2016) and can be heterogeneous in terms of virulence (Maury et al., 2016; Maury et al., 2017). Although with limited discriminatory power, *Lm* can also be distinguished into 13 different serotypes on the basis of the flagellar and somatic antigenic differences, that can be grouped into 6 PCR-based genoserogroups (IIa, IIb, IIc, IVb, IVb-v1, L) (Doumith et al., 2004; Leclercq et al., 2011). Together, serogroups IVb and IIb (lineage I), and IIa (lineage II), are responsible for a large part of confirmed cases of listeriosis (Lomonaco et al., 2015).

Listeriosis is a notifiable disease in North America and in many European countries (Camargo et al., 2017), but not in Brazil. Existing national epidemiological surveillance programs actively detect and investigate clusters of cases, identify the origin

of food contamination and implement control measures. Whole genome sequencing (WGS) has recently emerged as a powerful tool for national and international outbreak investigations (Schmid et al., 2014; Ruppitsch et al., 2015; Moura et al., 2017; Nielsen et al., 2017; Van Walle et al., 2018). WGS allows an unprecedented subtyping resolution by analysis of the core genome, as well as the accessory genome. Questions related to potential virulence heterogeneity and antimicrobial resistance genes can also be answered, and outbreak investigations have become more precise by correlating epidemiological data with genetic characteristics of the isolates involved (Nielsen et al., 2017; Lüth et al., 2018; Schürch et al., 2018).

Considering the importance of listeriosis, the widespread distribution of *Lm* and the absence of data from Brazil and South America, we used WGS to subtype and characterize *Lm* isolates recovered in the last decades from beef, food production environment (PE) and clinical samples from eight Brazilian states. This information was then compared with genomes from other countries in South America, North America and Europe.

Results and Discussion

In Brazil, few reports of listeriosis outbreaks or sporadic cases have been described and often no epidemiological link has been established (Camargo et al., 2017). In addition, isolates from outbreaks have only been superficially characterized. To better understand the population structure and molecular diversity of *Lm* in Brazil, we sequenced 35 genomes representative of isolation dates, geographic extent and sample origin. The quality metrics of the sequence data and draft assemblies are available in the **Supporting Information Table S1.** All draft genomes met the quality standards necessary for cgMLST analysis (Moura et al., 2016).

Lm *Population Structure*

The Brazilian isolates were grouped by lineage I (n=22; 63%) and lineage II (n=13; 37%) and by genoserogroups IVb (n=10; 28.6%), IIc (n=9; 25.7%), IIb (n=7; 20%), IVb-v1 (n=5; 14.3%), and IIa (n=4; 11.4%). On the basis of 7-loci MLST, isolates were distributed among 11 different sequence types (STs) (**Table 1**) and 10 clonal complexes (CCs), most of them previously reported in South America (**Fig. 1, Supporting Information Table S2**). ST1, ST2, and ST520 (serogroup IVb), ST218 (IVb-v1), ST3, ST288 (IIb), ST9 (IIc), and ST8 (IIa) had been previously reported in dairy industries and retail products in Brazil (Chenal-Francisque et al., 2011; Haase et al., 2014).

At the level of core genome, isolates were distributed among 10 sublineages (SLs) and represented 33 different cgMLST types (CTs) (**Fig. 2A**). For a better understanding of the broader phylogenetic context, the 35 Brazilian *Lm* isolates were included in a comparison of global data from 1,696 genomes (Moura et al., 2016) and 113 additional genomes from South America (**Supporting Information Table S3)**, from Chile (*n* = 95), Peru (*n* = 13), Colombia (*n* = 2), Brazil (*n* = 1), Venezuela (*n* = 1) and Argentina (*n* = 1). Consistent with MLST, the Brazilian isolates were mostly represented by sublineages that have been identified worldwide (**Fig. 3**). Out of 33 cgMLST types identified in this study, 31 were new in BIGSdb-*Lm* database and were assigned to CT4418-CT4434 and CT4436-CT4449 (**Fig. 2A**). Only L1-SL218-ST218-CT4015 and L2-SL7-ST12-CT720 types have been previous reported in Europe (Charlier et al., 2017; Maury et al., 2017) and in the US (SRA accession no. SRR1664806), suggesting a global intercontinental spread of these strains. The high proportion of new CTs (94%) found in this study highlights the current scarcity of genomes from South America in public databases

(currently mostly populated by genomes from North America, Europe and Australia) and the importance of more genomic studies to fully understand *Lm* biodiversity.

Long-Term Strain Persistence in Brazil

Interestingly, two CTs (highlighted in **Fig. 2A**) included more than one isolate (cgMLST similarity ≥ 99.6%), suggesting a possible epidemiological link (Moura et al., 2016). To confirm the close phylogenetic relationships observed using cgMLST (which covers 1748 genes, corresponding to ~1.6 Mb of *Lm* genome), isolates belonging to the same CT were also compared at the level of whole genome (~3.0 Mb, comprising all coding and non-coding regions) by single nucleotide polymorphisms analysis (wgSNP).

Isolates CLIST 2167 (collected in 1997, human source) and CLIST 3192 (collected in 2007 from raw beef), assigned as L1-SL315-ST520-CT4429, were collected a decade apart in São Paulo (SP) state and showed no allelic differences on the basis of cgMLST and wgSNP analyses.

The two isolates assigned to L2-SL9-ST9-CT4420 (isolate 19, collected in 2009 from meat handles and isolate 508, collected in 2012 from raw beef) were recovered 4 years apart from the same food processing plant in Minas Gerais (MG) state. These isolates differed by only 5 cgMLST alleles and 108 wgSNPs, the latter located in 12 genes (85 synonymous and 19 non-synonymous SNPs) and 4 non-coding regions. The high number of SNPs detected within only 12 genes suggest the occurrence of genetic recombination events. Whether these genetic modifications contributed to the adaptation or environmental persistence of these isolates remains unclear.

Despite the lack of trace back and/or forward investigations to confirm the epidemiological association between these isolate pairs, their high genetic relatedness highlights a possible long-term *Lm* contamination in these food production plants.

Antibiotic and stress resistance traits

None of the *Lm* isolates harbored any of the screened resistance genes towards disinfectants (**Fig. 2B**, **Supporting Information Table S4**). Intrinsic antibiotic resistance genes (*lmo0919, norB, lmo0441, sul, fosX*) were present in all isolates. Two isolates (cgMLST types L1-SL3-ST3-CT4445 and - L1-SL3-ST3-CT4446; **Fig. 2B**) exhibited a C(187)T mutation in the *fosX* gene (resistance to fosfomycin). This new mutation results in a premature stop codon (PMSC) at residue 63, likely resulting in complete susceptibility to fosfomycin (Scortti et al., 2018). Acquired antibiotic resistance traits were not detected, with the exception of *aacA4* (resistance to aminoglycosides) present in one isolate (isolate CLIST 2083, L1-SL3-ST3-CT4444).

The *Listeria* genomic island (LGI-2), which was previously identified in lineage I isolates and that is associated with arsenic resistance (Lee et al., 2017) was detected in 23% (8/35) of isolates from both lineages I and II (**Fig. 2B**). Nearly all lineage II isolates carried SSI-1 (**Fig. 2B**), whereas SSI-2 was only present in one isolate (CLIST 3864, L2-SL121-ST121-CT4428). These stress survival islets contributes to *Lm* survival and grow in adverse environmental conditions, such as low pH and high salt concentrations (Ryan et al., 2010) or alkaline and oxidative stress conditions (Harter et al., 2017), facilitating the adaptation and persistence in production environments. A similar SSI prevalence was previously reported in Chile (Toledo et al., 2018) ($P = 0.13$; Chi-square test), suggesting these mechanisms are typically present in *L. monocytogenes*, particularly in lineage II. Further studies will allow a better understanding as to whether these genes confer a selective advantage to the isolates described here.

Virulence Genetic Traits

As expected, the pathogenic island LIPI-1 was present in all isolates (**Fig. 2B**). Despite its high degree of conservation, two clinical isolates CLIST 2137 (L2-SL9-ST9-CT4425) and CLIST 2140 (L2-SL7-ST7-CT720) carried frameshift and mutations in the *prfA* gene leading to PMSCs (L(221)- and G(16)Stop, respectively) (**Fig. 2B**). The *prfA* gene regulates the transcription of several *Lm* genes expressed during infection (Chakraborty et al., 1992; Milohanic et al., 2003; de las Heras et al., 2011) and truncations in this gene can impact hemolysis and impair *Lm* virulence (Rupp et al., 2015; Maury et al., 2017). Results are in agreement with the previous reports that showed that the cgMLST type L2-SL7-ST7-CT720 is non-hemolytic due to a *prfA* PMSC (Maury et al., 2017).

Consistent with *Lm* phylogeny, isolates from lineage I also carried LIPI-3 and LIPI-4 islands. LIPI-3, which encodes listeriolysin S (Cotter et al., 2008; Quereda et al., 2016), was present in all lineage I isolates except those from sublineages SL2 and SL315. LIPI-4, which has been shown to increase *Lm* neuro- and placental tropism (Maury et al., 2016) was present only in 2 isolates (L1-SL315-ST520-CT4429; **Fig. 2**), consistent with previous reports of its presence in SL315 (Moura et al., 2016).

Alleles encoding InlA truncated variants, impaired in epithelial invasion (Lecuit et al., 1997; Lecuit et al., 2001), were identified in all SL121 and SL9 isolates, obtained from food or associated environments (**Fig. 2B**). The detected PMSCs included the types 6 (Q(492)Stop), 11 (W(685)Stop), and 19 (E(326)Stop), previously described (Olier et al., 2003; Rousseaux et al., 2004; Gelbíčová et al., 2015; Moura et al., 2016; Toledo et al., 2018). Interestingly, both cgMLST types involved in the persistent contaminations described above carried either deletions or mutations in the *inlA* gene. Isolates belonging to L1-SL315-ST520-CT4429 harbored a 3 amino acid deletion at residue 737 (pre-anchor

domain) in InlA internalin, whereas in L2-SL9-ST9-CT4420 isolates carried the PMSC at residue 326 (leucine-rich repeat motifs). Whether these favor strain persistence remains to be clarified.

Other genes important for *Lm* virulence and stress response also carried mutations or frameshifts leading to PMSCs (eg. *aut*, *oppA*, *agrC*, *mouR*) and/or were evenly distributed across the *Lm* phylogeny (eg. *inlG*, *bapL*) (**Fig. 2B).** Further studies will clarify the impact of these mutations on *Lm* virulence phenotypes.

Conclusions

Our results constitute a first contribution to understanding the genomic diversity and epidemiology of *Lm* in Brazil and South America. This study also highlights the importance of implementing WGS-based epidemiological surveillance programs to detect transmission chains and uncover long-term persistence of *Lm* in processing environments. Future work will clarify the impact of the genetic traits and mutations found in this study on *Lm* virulence and persistence.

Experimental procedures

Lm *Isolates Selection*

A collection of Brazilian isolates, most of them kindly provided by FioCruz (CLIST collection), were previously characterized (Camargo et al., 2015; Camargo et al., 2016). Based on this work, a total of 35 representative *Lm* isolates were selected for WGS based on their year of isolation (from 1978 to 2013), serotypes (1/2a, n = 7; 1/2b, n = 7; 1/2c, n = 6; 4b, n = 15), sources (PE, n = 5; beef, n = 16; and clinical, n = 14) and geographic distribution in the country (from eight states). The major serotypes from lineages I and II were included due to their epidemiological relevance (Hofer et al., 2006; Vallim et al.,

2015). Pure cultures were maintained at −80°C in brain heart infusion broth (BHI, Oxoid Ltd., Basingstoke, England) in the presence of 20% (v/v) glycerol (Merck, Whitehouse Station, NJ, USA), and recovered by growth on BHI agar prior use.

Genome Sequencing and Assembly

The genome sequence data was obtained after DNA extraction using the DNeasy Blood and Tissue kit (QIAGEN, Germany) following the protocol for purification of total DNA from Gram-positive bacteria, library preparation with Ovation Ultralow Library System V2 (Nugen, San Carlos, CA, USA) and sequencing by Illumina MiSeq platform. Paired-ends reads of 300 bp (Illumina, San Diego, CA, USA) included an average genome coverage of 169X (listed in **Supporting Information Table S1**). Reads were trimmed using Trimmomatic version 0.35 (Bolger et al., 2014) and assemblies were obtained using SPAdes version 3.10 (Anton et al., 2012).

In Silico Molecular Typing

PCR-serogroups (Doumith et al., 2004), MLST (Ragon et al., 2008) and cgMLST (Moura et al., 2016) profiles were extracted from draft assemblies as previously described (Moura et al., 2016). MLST profiles were compared against the 270 other public profiles from South America available at the Institute Pasteur MLST database (BIGSdb-*Lm*, http://bigsdb.pasteur.fr/listeria/), using the minimum spanning tree clustering method implemented in BioNumerics v7.6 (Applied-Maths).

SLs and cgMLST types (CTs) were assigned in BIGSdb-*Lm* database based on the cgMLST profiles and the allelic cut-offs previously established (150 allelic differences for SLs and 7 allelic differences from CTs; Moura et al. (2016)). New CTs were assigned to profiles not previously registered in BIGSdb-*Lm*.

Genomes were compared against the public collection of 1,696 *Lm* genomes collected mostly in Europe and North America (Moura et al., 2016) and 113 additional genomes from South America (listed in **Supporting Information Table S3**) retrieved from NCBI database (https://www.ncbi.nlm.nih.gov/; accessed on 01 Oct 2018). Dendrograms were built based on cgMLST profiles similarities in BioNumerics v.7.6 (http://www.applied-maths.com) using categorical differences and single-linkage clustering method. Trees were visualized by using iTol v.4.2 (Letunic and Bork, 2016).

Isolates belonging to the same cgMLST type were further analyzed at the level of whole-genome single nucleotide polymorphisms (wgSNP) by mapping their sequence reads against a representative reference assembly for each corresponding CT (isolates 19 and CLIST 2167, for CT4420 and CT4429, respectively) using the snippy pipeline (https://github.com/tseemann/snippy).

Virulence and Resistance Genotyping

The presence of 248 genes (listed in **Supporting Information Table S4**) previously identified as involved in stress and antibiotic resistance, biofilm formation and virulence (Chen et al., 2005; Kuenne et al., 2013; Gahan and Hill, 2014; Wattam et al., 2014; Maury et al., 2016) was accessed *in silico* in BIGSdb-*Lm* using BLASTN algorithm (Jolley and Maiden, 2010) as previously described (Moura et al., 2016). Individual gene alignments were performed using MUSCLE (Edgar, 2004).

Data Availability

The genomes obtained in this study were deposited in NCBI/EMBL/DDBJ databases under the project accession number PRJEB31124.

Acknowledgments

The authors express their gratitude to Deyse Vallim and Ernesto Hofer from the Laboratory of Bacterial Zoonoses (FioCruz, Brazil) for provision of bacterial strains from CLIST collection. The present material is based upon work supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

All authors designed the research; J.A and D.R.C. performed the genomic sequencing; N.H and S.S. assembled the genomes; A.M performed the genomic analyses; A.C.C. and A.M. wrote the paper; all authors edited the paper.

References

Anton, B., Sergey, N., Dmitry, A., Alexey A., G., Mikhail, D., Alexander S., K. et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455-477.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Camargo, A.C., Vallim, D.C., Hofer, E., and Nero, L.A. (2016) Molecular serogrouping of *Listeria monocytogenes* from Brazil using PCR. *Journal of Food Protection* **79**: 144-147.

Camargo, A.C., Woodward, J.J., Call, D.R., and Nero, L.A. (2017) *Listeria monocytogenes* in food-processing facilities, food contamination, and human listeriosis: the Brazilian scenario. *Foodborne Pathogens and Disease* **14**: 623-636.

Camargo, A.C., Castilho, N.P.A., Silva, D.A.L., Vallim, D.C., Hofer, E., and Nero, L.A. (2015) Antibiotic resistance of *Listeria monocytogenes* isolated from meat-processing environments, beef products, and clinical cases in Brazil. *Microbial Drug Resistance* **21**: 458-462.

Chakraborty, T., Leimeister-Wächter, M., Domann, E., Hartl, M., Goebel, W., Nichterlein, T., and Notermans, S. (1992) Coordinate regulation of virulence genes in *Listeria monocytogenes* requires the product of the *prfA* gene. *Journal of Bacteriology* **174**: 568-574.

Charlier, C., Perrodeau, É., Leclercq, A., Cazenave, B., Pilmis, B., Henry, B. et al. (2017) Clinical features and prognostic factors of listeriosis: the MONALISA national prospective cohort study. *The Lancet Infectious Diseases* **17**: 510-519.

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., and Jin, Q. (2005) VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Research* **33**: D325-D328.

Chenal-Francisque, V., Lopez, J., Cantinelli, T., Caro, V., Tran, C., Leclercq, A. et al. (2011) Worldwide distribution of major clones of *Listeria monocytogenes*. *Emerging Infectious Diseases* **17**: 1110-1112.

Cotter, P.D., Draper, L.A., Lawton, E.M., Daly, K.M., Groeger, D.S., Casey, P.G. et al. (2008) Listeriolysin S, a Novel Peptide Haemolysin Associated with a Subset of Lineage I *Listeria monocytogenes*. *PLOS Pathogens* **4**: e1000144.

de las Heras, A., Cain, R.J., Bielecka, M.K., and Vázquez-Boland, J.A. (2011) Regulation of Listeria virulence: PrfA master and commander. *Current Opinion in Microbiology* **14**: 118-127.

de Noordhout, C.M., Devleesschauwer, B., Angulo, F.J., Verbeke, G., Haagsma, J., Kirk, M. et al. (2014) The global burden of listeriosis: a systematic review and meta-analysis. *The Lancet Infectious Diseases* **14**: 1073-1082.

Doumith, M., Buchrieser, C., Glaser, P., Jacquet, C., and Martin, P. (2004) Differentiation of the major *Listeria monocytogenes* serovars by multiplex PCR. *Journal of Clinical Microbiology* **42**: 3819-3822.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Gahan, C., and Hill, C. (2014) *Listeria monocytogenes:* survival and adaptation in the gastrointestinal tract. *Frontiers in Cellular and Infection Microbiology* **4**.

Gelbíčová, T., Koláčková, I., Pantůček, R., and Karpíšková, R. (2015) A novel mutation leading to a premature stop codon in inlA of *Listeria monocytogenes* isolated from neonatal listeriosis. *The New Microbiologica* **38**: 293-296.

Haase, J.K., Didelot, X., Lecuit, M., Korkeala, H., and Achtman, M. (2014) The ubiquitous nature of *Listeria monocytogenes* clones: a large-scale multilocus sequence typing study. *Environmental Microbiology* **16**: 405-416.
Harter, E., Wagner, E.M., Zaiser, A., Halecker, S., Wagner, M., and Rychli, K. (2017) The novel stress survival islet 2 (SSI-2), predominantly present in *Listeria monocytogenes* strains of ST121, is involved in alkaline and oxidative stress response. *Applied and Environmental Microbiology*.
Hofer, E., Reis, C.M.F.d., and Hofer, C.B. (2006) Sorovares de *Listeria monocytogenes* e espécies relacionadas, isoladas de material clínico humano. *Revista da Sociedade Brasileira de Medicina Tropical* **39**: 32-37.
Jolley, K.A., and Maiden, M.C. (2010) BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**: 595.
Kuenne, C., Billion, A., Mraheil, M.A., Strittmatter, A., Daniel, R., Goesmann, A. et al. (2013) Reassessment of the *Listeria monocytogenes* pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. *BMC Genomics* **14**: 47.
Leclercq, A., Chenal-Francisque, V., Dieye, H., Cantinelli, T., Drali, R., Brisse, S., and Lecuit, M. (2011) Characterization of the novel *Listeria monocytogenes* PCR serogrouping profile IVb-v1. *International Journal of Food Microbiology* **147**: 74-77.
Lecuit, M., Ohayon, H., Braun, L., Mengaud, J., and Cossart, P. (1997) Internalin of *Listeria monocytogenes* with an intact leucine-rich repeat region is sufficient to promote internalization. *Infection and Immunity* **65**: 5309-5319.
Lecuit, M., Vandormael-Pournin, S., Lefort, J., Huerre, M., Gounon, P., Dupuy, C. et al. (2001) A transgenic model for listeriosis: role of internalin in crossing the intestinal barrier. *Science* **292**: 1722-1725.
Lee, S., Ward, T.J., Jima, D.D., Parsons, C., and Kathariou, S. (2017) The arsenic resistance Listeria genomic island LGI2 exhibits sequence and integration site diversity and propensity for three *Listeria monocytogene* clones with enhanced virulence. *Applied and Environmental Microbiology* **83**: e01189-01117.
Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* **44**: W242-W245.
Lomonaco, S., Nucera, D., and Filipello, V. (2015) The evolution and epidemiology of *Listeria monocytogenes* in Europe and the United States. *Infection, Genetics and Evolution* **35**: 172-183.
Lüth, S., Kleta, S., and Al Dahouk, S. (2018) Whole genome sequencing as a typing tool for foodborne pathogens like *Listeria monocytogenes* – the way towards global harmonisation and data exchange. *Trends in Food Science & Technology* **73**: 67-75.
Maury, M.M., Chenal-Francisque, V., Bracq-Dieye, H., Han, L., Leclercq, A., Vales, G. et al. (2017) Spontaneous loss of virulence in natural populations of *Listeria monocytogenes*. *Infection and Immunity* **85**: e00541-00517.
Maury, M.M., Tsai, Y.-H., Charlier, C., Touchon, M., Chenal-Francisque, V., Leclercq, A. et al. (2016) Uncovering *Listeria monocytogenes* hypervirulence by harnessing its biodiversity. *Nat Genet* **48**: 308-313.
Milohanic, E., Glaser, P., Coppée, J.-Y., Frangeul, L., Vega, Y., Vázquez-Boland, J.A. et al. (2003) Transcriptome analysis of *Listeria monocytogenes* identifies three groups of genes differently regulated by PrfA. *Molecular Microbiology* **47**: 1613-1625.
Moura, A., Tourdjman, M., Leclercq, A., Hamelin, E., Laurent, E., Fredriksen, N. et al. (2017) Real-time whole-genome sequencing for surveillance of *Listeria monocytogenes*, France. *Emerging Infectious Diseases* **23**: 1462-1470.

Moura, A., Criscuolo, A., Pouseele, H., Maury, M.M., Leclercq, A., Tarr, C. et al. (2016) Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. *Nature Microbiology* **2**: 16185.
Nielsen, E.M., Björkman, J.T., Kiil, K., Grant, K., Dallman, T., Painset, A. et al. (2017) Closing gaps for performing a risk assessment on *Listeria monocytogenes* in ready-to-eat (RTE) foods: activity 3, the comparison of isolates from different compartments along the food chain, and from humans using whole genome sequencing (WGS) analysis. *EFSA Supporting Publications* **14**: 1151E.
Olier, M., Pierre, F., Rousseaux, S., Lemaître, J.-P., Rousset, A., Piveteau, P., and Guzzo, J. (2003) Expression of Truncated Internalin A Is Involved in Impaired Internalization of Some *Listeria monocytogenes* Isolates Carried Asymptomatically by Humans. *Infection and Immunity* **71**: 1217-1224.
Orsi, R.H., Bakker, H.C.d., and Wiedmann, M. (2011) *Listeria monocytogenes* lineages: Genomics, evolution, ecology, and phenotypic characteristics. *International Journal of Medical Microbiology* **301**: 79-96.
Quereda, J.J., Dussurget, O., Nahori, M.-A., Ghozlane, A., Volant, S., Dillies, M.-A. et al. (2016) Bacteriocin from epidemic Listeria strains alters the host intestinal microbiota to favor infection. *Proceedings of the National Academy of Sciences* **113**: 5706-5711.
Ragon, M., Wirth, T., Hollandt, F., Lavenir, R., Lecuit, M., Le Monnier, A., and Brisse, S. (2008) A new perspective on *Listeria monocytogenes* Evolution. *PLOS Pathogens* **4**: e1000146.
Rousseaux, S., Olier, M., Lemaître, J.P., Piveteau, P., and Guzzo, J. (2004) Use of PCR-Restriction Fragment Length Polymorphism of *inlA* for Rapid Screening of *Listeria monocytogenes* Strains Deficient in the Ability To Invade Caco-2 Cells. *Applied and Environmental Microbiology* **70**: 2180-2185.
Rupp, S., Aguilar-Bultet, L., Jagannathan, V., Guldimann, C., Drögemüller, C., Pfarrer, C. et al. (2015) A naturally occurring prfA truncation in a *Listeria monocytogenes* field strain contributes to reduced replication and cell-to-cell spread. *Veterinary Microbiology* **179**: 91-101.
Ruppitsch, W., Pietzka, A., Prior, K., Bletz, S., Fernandez, H.L., Allerberger, F. et al. (2015) Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of *Listeria monocytogenes*. *Journal of Clinical Microbiology* **53**: 2869-2876.
Ryan, S., Begley, M., Hill, C., and Gahan, C.G.M. (2010) A five-gene stress survival islet (SSI-1) that contributes to the growth of *Listeria monocytogenes* in suboptimal conditions. *Journal of Applied Microbiology* **109**: 984-995.
Schmid, D., Allerberger, F., Huhulescu, S., Pietzka, A., Amar, C., Kleta, S. et al. (2014) Whole genome sequencing as a tool to investigate a cluster of seven cases of listeriosis in Austria and Germany, 2011–2013. *Clinical Microbiology and Infection* **20**: 431-436.
Schürch, A.C., Arredondo-Alonso, S., Willems, R.J.L., and Goering, R.V. (2018) Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. *Clinical Microbiology and Infection* **24**: 350-354.
Scortti, M., Han, L., Alvarez, S., Leclercq, A., Moura, A., Lecuit, M., and Vazquez-Boland, J. (2018) Epistatic control of intrinsic resistance by virulence genes in *Listeria*. *PLOS Genetics* **14**: e1007525.
Silk, B.J., Date, K.A., Jackson, K.A., Pouillot, R., Holt, K.G., Graves, L.M. et al. (2012) Invasive listeriosis in the foodborne diseases active surveillance network (foodNet), 2004–2009: further targeted prevention needed for higher-risk groups. *Clinical Infectious Diseases* **54**: S396-S404.

Toledo, V., den Bakker, H.C., Hormazábal, J.C., González-Rocha, G., Bello-Toledo, H., Toro, M., and Moreno-Switt, A.I. (2018) Genomic diversity of *Listeria monocytogenesis* isolated from clinical and non-clinical samples in Chile. *Genes* **9**: 1-12.
Vallim, D.C., Barroso Hofer, C., Lisba, R.d.C., Victor Barbosa, A., Alves Rusak, L., Reis, C.M.F.d., and Hofer, E. (2015) Twenty years of Listeria in Brazil: occurrence of Listeria species and *Listeria monocytogenes* serovars in food samples in Brazil between 1990 and 2012. *BioMed Research International* **2015**: 1-8.
Van Walle, I., Björkman, J.T., Cormican, M., Dallman, T., Mossong, J., Moura, A. et al. (2018) Retrospective validation of whole genome sequencing-enhanced surveillance of listeriosis in Europe, 2010 to 2015. *Eurosurveillance* **23**: 1700798.
Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L. et al. (2014) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research* **42**: D581-D591.

Table 1. Genomic typing of *Listeria monocytogenes* isolates recovered in Brazil from 1978 to 2013.

Isolate	Source[1]	State[2]	Year	Lineage	Serotype	Serogroup[3]	CC (MLST)[4]	ST (MLST)[4]	SL (cgMLST)[5]	CT (cgMLST)[5]	Accession numbers
CLIST 2140	Human - Placenta	PE	1978	II	1/2a	IIa	CC7	ST12	SL7	CT720	ERR3199887
CLIST 732	Raw beef	MT	2009	II	1/2a	IIa	CC8	ST8	SL8	CT4418	ERR3199900
CLIST 3727	Raw beef	SP	2003	II	1/2a	IIa	CC8	ST1419	SL8	CT4419	ERR3199902
CLIST 3864	Raw beef	RS	2002	II	1/2a	IIa	CC121	ST121	SL121	CT4428	ERR3199903
19	FPE	MG	2009	II	1/2c	IIc	CC9	ST9	SL9	CT4420	ERR3199909
45	Raw beef	MG	2010	II	1/2c	IIc	CC9	ST9	SL9	CT4424	ERR3199911
227	FPE	RS	2010	II	1/2c	IIc	CC9	ST9	SL9	CT4422	ERR3199910
508	Raw beef	MG	2012	II	1/2c	IIc	CC9	ST9	SL9	CT4420	ERR3199912
581	Raw beef	MG	2012	II	1/2c	IIc	CC9	ST9	SL9	CT4423	ERR3199913
CLIST 2137	Human- CSF	PR	1983	II	1/2a	IIc	CC9	ST9	SL9	CT4425	ERR3199886
CLIST 3186	Raw beef	SP	2006	II	1/2a	IIc	CC9	ST9	SL9	CT4426	ERR3199901
CLIST 3726	Raw beef	SP	2003	II	1/2c	IIc	CC9	ST9	SL9	CT4421	ERR3199914
CLIST 3732	Human - CSF	DF	1989	II	1/2a	IIc	CC9	ST9	SL9	CT4427	ERR3199888
7	Raw beef	MG	2009	I	1/2b	IIb	CC3	ST3	SL3	CT4448	ERR3199904
CLIST 441	Raw beef	MT	2010	I	1/2b	IIb	CC3	ST3	SL3	CT4447	ERR3199905
CLIST 2083	Human - Blood	RJ	2008	I	1/2b	IIb	CC3	ST3	SL3	CT4444	ERR3199889
CLIST 2138	Human - CSF	SP	1982	I	1/2b	IIb	CC3	ST3	SL3	CT4445	ERR3199890
CLIST 3865	Raw beef	SP	2002	I	1/2b	IIb	CC3	ST3	SL3	CT4446	ERR3199906
CLIST 3869	Raw beef	RJ	2004	I	1/2b	IIb	CC3	ST3	SL3	CT4449	ERR3199907
CLIST 3870	Raw beef	MT	2004	I	1/2b	IIb	CC288	ST288	SL288	CT4439	ERR3199908
CLIST 1011	Human - Blood	SP	1985	I	4b	IVb	CC1	ST1	SL1	CT4437	ERR3199891
CLIST 1015	Human - Blood	SP	1985	I	4b	IVb	CC1	ST1	SL1	CT4438	ERR3199892
CLIST 3735	Human - Blood	SP	1985	I	4b	IVb	CC1	ST1	SL1	CT4436	ERR3199898
233	FPE	MG	2012	I	4b	IVb	CC2	ST2	SL2	CT4431	ERR3199915
CLIST 3723	Human - CSF	RJ	1990	I	4b	IVb	CC2	ST2	SL2	CT4433	ERR3199896
CLIST 3724	Raw beef	RJ	2003	I	4b	IVb	CC2	ST2	SL2	CT4434	ERR3199920
CLIST 3729	Animal	RS	1991	I	4b	IVb	CC2	ST2	SL2	CT4430	ERR3199917
CLIST 3739	Human - CSF	PR	2000	I	4b	IVb	CC2	ST2	SL2	CT4432	ERR3199899
CLIST 3734	Human - CSF	PE	1989	I	4b	IVb-v1	CC218	ST218	SL218	CT4015	ERR3199897
CLIST 2930	Human - Blood	RJ	2004	I	4b	IVb-v1	CC218	ST218	SL218	CT4440	ERR3199895
CLIST 2168	Human - CSF	RJ	2007	I	4b	IVb-v1	CC218	ST218	SL218	CT4441	ERR3199894
1282	FPE	SP	2013	I	4b	IVb-v1	CC218	ST218	SL218	CT4442	ERR3199916
74	Raw beef	MG	2010	I	4b	IVb-v1	CC218	ST218	SL218	CT4443	ERR3199918
CLIST 2167	Human - CSF	SP	1997	I	4b	IVb	CC315	ST520	SL315	CT4429	ERR3199893
CLIST 3192	Raw beef	SP	2006	I	4b	IVb	CC315	ST520	SL315	CT4429	ERR3199919

[1] FPE, food processing environment; CSF, cerebrospinal fluid; [2] Brazilian state: PE, Pernambuco; MT, Mato Grosso; SP, São Paulo; RS, Rio Grande do Sul; MG, Minas Gerais; PR, Paraná; DF, Distrito Federal; RJ, Rio de Janeiro; [3] according to Doumith et al. (2004) and Leclercq et al. (2011); [4] clonal complex (CC) and sequence type (ST) defined according to Ragon et al. (2008); [5] sublineage (SL) and cgMLST type (CT) defined according to Moura et al. (2016).

Figure Legends

Figure 1. Minimum spanning tree based on the 7-loci MLST profiles of the 35 isolates included in this study together with the 270 profiles from South America publicly available at BIGSdb-Lm (listed in Table S2). Sequence types (STs) are represented by colored circles where size is proportional to the number of isolates. Grey zones denote MLST clonal complexes (CCs). The number of allelic differences between profiles are indicated on the branches.

Figure 2. Resistance and virulence gene profiles of the 35 isolates sequenced in this study. A) Single-linkage clustering based on the cgMLST profiles. Branches are coloured by phylogenetic lineage (L1, red; L2, orange) and labelled by sublineage (SL). Information on the BIGSdb identifier, isolate's name, origin (H, human; F, food; PE, processing environment), state (PE, Pernambuco; MT, Mato Grosso; SP, São Paulo; RS, Rio Grande do Sul; MG, Minas Gerais; PR, Paraná; DF, Distrito Federal; and RJ, Rio de Janeiro), year of isolation, serotype (serogroup) and cgMLST type are provided. Colour-filled boxes represent the presence of the different genetic traits. Empty boxes represent genes with truncations leading to premature stop codons. Absent genes are marked in white.

Figure 3. Phylogenetic placement of the 35 isolates included in this study together with the worldwide collection of 1,696 genomes (Moura et al., 2016) and 113 additional genomes from South America retrieved from NCBI database (last accessed 01-10-2018). The dendrogram was obtained using the single-linkage method on the cgMLST allelic profiles of the 1844 isolates. Branches are coloured by phylogenetic lineage (L1, red; L2, orange; L3, green; L4, blue).

Sublineages (SL) with more than 30 isolates are labelled in the tree. Inner rings show the isolate's source type and geographic location, according to the colour codes shown on the left. Isolates from this study are marked in blue in the most external ring.